Lecture 35

In this lecture we continue our study of statistical distributions, looking at the standard deviation.

Standard Deviation

The positive square root of the variance is called the standard deviation, denoted by σ or σ_X.

(σ: lower case Greek letter "sigma")

So $Var(X) = \sigma_X^2$.

For example, in our last lecture we found the expectation and variance for the score X when a die is cast. The results were:

$$E(X) = \frac{7}{2} \quad \text{and} \quad Var(X) = \frac{35}{12}$$

So we can write $\sigma_X^2 = \frac{35}{12}$

whence $\sigma_X = \sqrt{\frac{35}{12}} \simeq 1.708.$

The standard deviation is often used as a measure of spread.

For example, it's common to ask a question like this: How many standard deviations away from the mean was this particular observation?

For most common distributions, nearly all observations are within 3 standard deviations on either side of the mean.

To illustrate this, in the last example $E(X) - 3\sigma_X \approx -1.62$ and $E(X) + 3\sigma_X \approx 8.62$. The six possible observations when a die is cast are 1, 2, 3, 4, 5 and 6, which are all within the range from -1.62 to 8.62.

The binomial situation

Suppose that an experiment has probability p of success, and that it is carried out n times. Let X be the number of successes.

It can be shown that

$$E(X) = np$$

and

$$Var(X) = npq \quad = np(1-p)$$

so that

$$\sigma_X = \sqrt{npq}\,.$$

For example, if a fair coin is tossed 5 times then:

$$E(X) = 5\left(\tfrac{1}{2}\right) = 5/2 \quad = 2\tfrac{1}{2}$$

$$Var(X) = 5\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = 5/4 \quad = 1\tfrac{1}{4}$$

$$\sigma_X = \sqrt{5/4} = \sqrt{5}/2 \quad \simeq 1.118.$$

Population versus Sample

The values $E(x)$, $Var(x)$ and σ_x are *population parameters* when obtained from a probability distribution.

But each time we do an experiment we get *sample* values. This sample has its own mean, variance and standard deviation. Sometimes they will differ quite a bit from the corresponding population parameters.

The measures obtained from a sample are called *sample statistics*.

This ends the course.